

ANÁLISIS *in silico* DEL EFECTO DEL KNOCKOUT DE GENES EN EL CRECIMIENTO DE *Rhodospiridium toruloides*.

Castañeda, M. Teresita^{1,2,3✉}; Nuñez, Sebastián¹; Voget, Claudio³; De Battista, Hernán¹.

¹Grupo de Control Aplicado (GCA), Instituto LEICI, UNLP-CONICET, Facultad de Ingeniería, Universidad Nacional de La Plata, Argentina.

² Centro de Investigación y Desarrollo en Ciencias y Tecnología de Materiales (CITEMA), UTN-CICpBA, Laboratorio de Biotecnología, Departamento de Ingeniería Química, Universidad Tecnológica Nacional-Facultad Regional La Plata, Argentina

³ Centro de Investigación y Desarrollo en Fermentaciones Industriales (CINDEFI), UNLP-CONICET, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina.
Calle 47 y 115, (B1900ASH), La Plata, Argentina. Tel.: +54 221 483-3794.
e-mail: castaneda@biotec.quimica.unlp.edu.ar

Palabras clave: lípidos microbianos, biodiesel, *Rhodospiridium toruloides*, Análisis de balance de flujos, *in silico* knockout, esencialidad de los genes.

1.-INTRODUCCIÓN

El empleo de lípidos microbianos para la producción de biodiesel ha despertado interés en los últimos años como alternativa a los aceites vegetales, ya que presentan tiempos de producción considerablemente menores y son independientes de la localización geográfica y de las condiciones climáticas. Además, son fácilmente escalables y su cultivo es más económico [1, 2]. Estos lípidos se obtienen biotecnológicamente mediante el cultivo de microorganismos oleaginosos, los cuales disponen de la maquinaria metabólica necesaria para acumular más del 20% de su peso seco en lípidos, principalmente en forma de triglicéridos (TAG) [3]. A pesar de que existe una amplia gama de microorganismos oleaginosos, la levadura *Rhodospiridium toruloides* es una de las cepas más estudiadas para su empleo a escala industrial dada su versatilidad nutricional y elevados rendimientos lipídicos [4, 5]. En la cepa *wild-type* (WT), estos rendimientos son extremadamente variables y están mayoritariamente sujetos a parámetros relacionados con su cultivo [6]. Es por esto que los esfuerzos se concentran en la optimización de la producción de lípidos en la cepa nativa [5,7,8, 9] y en el mejoramiento de la cepa mediante ingeniería genética para la sobreproducción de lípidos [10, 11, 12].

En lo que refiere al mejoramiento genético para la sobreproducción de un metabolito de interés, una de las técnicas abordadas es el knockout de genes [13]. En términos sencillos, esta técnica consiste en “inactivar” ciertos genes específicos para silenciar determinadas reacciones metabólicas de modo de obligar al microorganismo a utilizar aquellas rutas metabólicas asociadas a la producción del metabolito de interés. Es fundamental para ello conocer de antemano la esencialidad de los genes involucrados en las reacciones, ya que el knockout de un gen esencial para el crecimiento resultará letal para el microorganismo y, por ende, no debe estar sujeto a modificación genética. Experimentalmente, la búsqueda de genes esenciales se lleva a cabo mediante técnicas aleatorias que resultan muy costosas y requieren mucho tiempo de implementación. La ingeniería metabólica mediante el Análisis de Balance de Flujos (FBA) ofrece una alternativa rápida y sencilla a los métodos experimentales para determinar, mediante simulación computacional, la esencialidad de los genes que intervienen en el metabolismo de un determinado microorganismo [14]. Estas técnicas hacen uso de modelos en los cuales están representadas las rutas metabólicas que describen el metabolismo de un microorganismo y con ello su potencialidad como fábricas celulares. Nuestro grupo de trabajo ha desarrollado recientemente el primer modelo curado a baja escala para la producción de lípidos en *R. toruloides* [6]. Sobre la base de este modelo, el objetivo de este trabajo es determinar mediante FBA la influencia del knockout de

genes sobre el crecimiento de *R. toruloides* utilizando glucosa y glicerol individualmente como fuentes de carbono y energía (FCE).

2.-MODELO Y MÉTODO DE ANÁLISIS

2.1.- Modelo metabólico de *R. toruloides* para la producción de lípidos.

Como se ha mencionado anteriormente, para el análisis de la esencialidad de genes en *R. toruloides* mediante FBA, se hizo uso de un modelo metabólico a baja escala recientemente publicado por nuestro grupo de trabajo [6]. Dicho modelo está compuesto por 93 metabolitos (m) y 104 reacciones (n) las cuales describen las principales rutas metabólicas: glicólisis, ruta de pentosas fosfato (PPP), ciclo de Krebs, ciclo de glioxilato y metabolismo central del nitrógeno. Además, el modelo se compone de un compartimento citosólico ($[c]$) y un compartimento mitocondrial ($[m]$) los cuales se conectan entre sí y con el medio extracelular ($[e]$) mediante reacciones de transporte de membrana. En este modelo, la producción de lípidos no-estructurales está representada mediante la síntesis de ácido palmítico ($C16$) y su posterior esterificación con glicerol para formar tripalmitina ($C51$), como TAG modelo. Por otra parte, la biomasa está representada por una pseudoecuación ($Cell_mass_Eqn$) que resulta de la combinación lineal de precursores metabólicos. Ambas reacciones tienen sus correspondientes reacciones de intercambio con el medio extracelular dadas por la reacción $TAGEx$ y $GrowthEx$, respectivamente. Si bien el modelo original contempla la utilización de cuatro FCE (glucosa, glicerol, xilosa y arabinosa) en este estudio nos centramos en las dos primeras. Por un lado, la glucosa es la FCE por excelencia, mientras que el glicerol es una FCE muy atractiva ya que constituye por sí mismo un subproducto de la reacción de transesterificación de TAG para la producción de biodiesel [7, 15].

Sobre la base de este modelo estequiométrico, se incorporaron 116 genes junto con las asociaciones gen-proteína-reacción (GPR), las cuales indican cómo están asociadas las reacciones con las proteínas del modelo (enzimas) y dichas proteínas con uno o más de los genes incorporados. Tanto los genes como las asociaciones GPR se obtuvieron a partir de bases de datos como UniProtKB [16].

2.2.- Simulación del knockout de genes mediante FBA.

El mapa metabólico anteriormente descrito está conformado por m metabolitos, cuya dinámica está dada por:

$$\frac{dc}{dt} = Sv - \mu c \quad (1)$$

donde S es la matriz estequiométrica con dimensiones $m \times n$, $c = (c_1 \ c_2 \ \dots \ c_m)^T$ es el vector con las concentraciones de los m metabolitos, $v = (v_1 \ v_2 \ \dots \ v_n)^T$ es el vector de n flujos intracelulares y μ es la velocidad específica de crecimiento [17]. En sistemas biológicos, la dinámica intracelular es mucho más rápida que la extracelular y el término de dilución ($-\mu c$) es despreciable frente a los valores de los flujos de las reacciones donde intervienen estos metabolitos. Luego, en estado estacionario:

$$Sv = 0 \quad (2)$$

Nuestro modelo, así como la mayoría de los modelos metabólicos, es indeterminado ($n > m$), con lo cual para lograr describir y predecir el comportamiento microbiano se requiere de técnicas de resolución que permitan acotar el espacio de soluciones.

FBA es una técnica predictiva que permite acotar el espacio de soluciones posibles mediante la optimización de una función objetivo elegida racionalmente, sujeto a restricciones metabólicas [18].

$$\begin{aligned} \text{máx.} \quad & Z = wv, \\ \text{sujeto a} \quad & Sv = 0, \\ & v_{\min} \leq v \leq v_{\max} \end{aligned} \quad (3)$$

donde w es el vector de coeficientes que define a la función objetivo (Z) como una combinación lineal de los flujos v_i . Los vectores v_{\min} y v_{\max} representan el límite inferior y superior para cada una de las reacciones. Los flujos de reacciones de intercambio con el

medio (biomasa, TAG, sustratos, O₂, CO₂, etc) son por definición igual a su velocidad específica de producción o consumo (q_i), en unidades de mmol/g h.

FBA se basa en la idea de que el comportamiento microbiano es restricto por naturaleza ya sea por termodinámica o capacidad máxima de enzimas o transportadores (restricciones no ajustables), o bien por restricciones cinéticas o regulatorias (restricciones ajustables). Además, esta técnica asume que el microorganismo evolutivamente persigue un objetivo condicionado por el entorno. Se han obtenido muy buenas predicciones empleando la reacción que describe la producción de biomasa como función objetivo [17].

Mediante FBA, se puede simular el efecto del knockout de un gen en el crecimiento utilizando la reacción de producción de biomasa como función objetivo y fijando los límites inferiores y superiores de la reacción o reacciones asociadas al gen en cero [19]. Para llevar a cabo estas simulaciones abarcando todos los genes incorporados al modelo se utilizó el Toolbox COBRA [20] para MATLAB, el cual contiene la función `deleteModelGenes`. Esta función, hace uso de las asociaciones GPR, las cuales utilizan las leyes de Boole para dictaminar como están conectadas las reacciones del modelo con los genes. Así, por ejemplo, cuando una reacción está catalizada por isozimas (dos enzimas diferentes catalizan la misma reacción), la asociación GPR contiene una regla “or”, lo que implica que puede silenciarse un gen sin afectar la reacción siempre y cuando haya un gen activo relacionado con una de las isozimas. Por otro lado, para reacciones catalizadas por enzimas con múltiples subunidades la asociación GPR contiene una regla “and” e implica que es suficiente con silenciar uno de los genes intervinientes la reacción, para que la misma tenga flujo cero. Finalmente, para reacciones catalizadas por enzimas codificadas por un único gen la asociación GPR no hace uso de reglas Boole y la delección del gen implica silenciar la reacción [18]. Una vez que la función `deleteModelGenes` relacionó los genes con las reacciones del modelo mediante las asociaciones GPR, lleva a cabo el knockout individual o de a pares de genes y mediante FBA predice el fenotipo de las cepas mutantes (KO).

2.3.- Análisis de la esencialidad de genes.

Como resultado de los knockouts individuales o de a pares de genes, el espacio de soluciones resulta restringido y la relación entre el flujo de la función objetivo (en este caso μ) para cepa KO respecto del valor para la cepa WT ($grRatio$) variará de acuerdo con la esencialidad de dicho gen.

$$grRatio = \frac{grRateKO}{grRateWT} \quad (4)$$

De acuerdo al valor del $grRatio$ se pueden dar tres situaciones [19]: (1) $grRatio=1$ implica que la remoción del gen no afecta el crecimiento del microorganismo (genes no esenciales), (2) $0 < grRatio < 1$ significa que la remoción del gen afecta al crecimiento pero la cepa aún puede adaptarse a crecer a un μ menor al μ óptimo y (3) $grRatio=0$ implica que la remoción del gen resulta letal para el microorganismo (gen esencial).

Independientemente de la clasificación anterior, se ha observado que la esencialidad de los genes del metabolismo es altamente dependiente del entorno, especialmente de la calidad del medio de cultivo. En base a esto, Zhang y Ren, 2015 [21] propusieron una clasificación más amplia de los genes esenciales. De acuerdo con esta clasificación, se conocen como genes condicionales esenciales a aquellos genes que son esenciales solo bajo ciertas circunstancias o condiciones de cultivo mientras que genes absolutamente esenciales o genes mínimos son aquellos que son necesarios para asegurar la viabilidad del microorganismo en un entorno con los nutrientes mínimos necesarios para su desarrollo. Por otra parte, esta clasificación habla también de genes esenciales redundantes cuya delección no produce por sí misma la letalidad, pero sí en combinación con otro gen involucrado en una/s reacción/es esencial/es del metabolismo.

Para el análisis de la esencialidad de los genes involucrados en el modelo se simuló el knockout de cada gen individualmente y el doble knockout de todas las combinaciones posibles de genes incorporados al modelo, empleando glucosa y glicerol como únicas FCE.

3.-RESULTADOS Y DISCUSIÓN

3.1.- Simulación del knockout individual de genes.

Por medio de FBA se determinó el efecto del knockout individual de cada gen incorporado al modelo metabólico de *R. toruloides* con glucosa y alternativamente con glicerol como única FCE. Como puede observarse en la Fig. 1A. de los 116 genes del modelo, 25 resultaron esenciales con glucosa como única FCE (eje abscisas). Del resto, las delecciones individuales de 19 genes tuvieron el efecto de disminuir la tasa de crecimiento (menor tonalidad) mientras que 72 de ellos resultaron no esenciales para el crecimiento ya que la tasa de crecimiento no se vio afectada tras el knockout de estos genes respecto a la cepa WT (mayor tonalidad). En lo que respecta al glicerol (Fig. 1B), el knockout individual de 25 genes resultó letal para la cepa (eje abscisas), mientras que la tasa de crecimiento fue menor a la WT como resultado del knockout individual de 17 genes (menor tonalidad) y 74 genes resultaron no esenciales, o sea, sus respectivas delecciones no afectaron el crecimiento (mayor tonalidad). En ambas FCE, a excepción de los 25 genes esenciales, el resto de los genes pueden ser sometidos a modificaciones genéticas con el objetivo de aumentar la tasa de producción de TAG. En cuanto a los genes que resultaron esenciales para el crecimiento de *R. toruloides* en cada FCE, se determinó la naturaleza de su esencialidad. De los 25 genes esenciales para cada FCE, 21 resultaron genes esenciales mínimos, o sea, son necesarios en condiciones óptimas de cultivo independientemente de la FCE empleada. Estos genes están asociados a enzimas que catalizan reacciones fundamentales de la glicólisis, de la ruta de pentosa fosfato, del ciclo de Krebs, de asimilación de sulfato y amonio, de fosforilación oxidativa, transportadores mitocondriales, bomba de protones, etc. Los 4 genes restantes resultaron ser genes condicionales esenciales, esto es, dependiendo del tipo de FCE resultaron esenciales para la producción de intermediarios metabólicos que intervienen en la síntesis de biomasa.

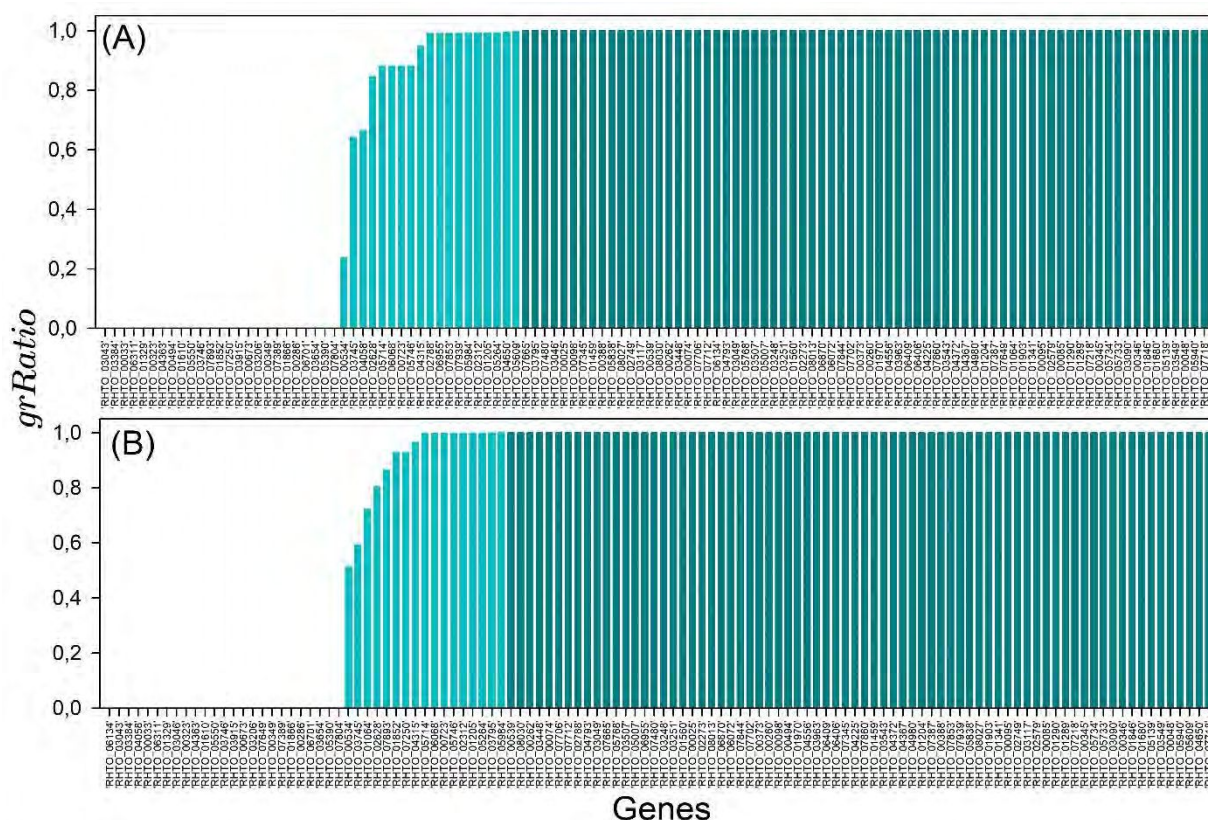


Fig. 1. Influencia del knockout individual de genes en el crecimiento de *R. toruloides* en glucosa (A) y glicerol (B) como FCE.

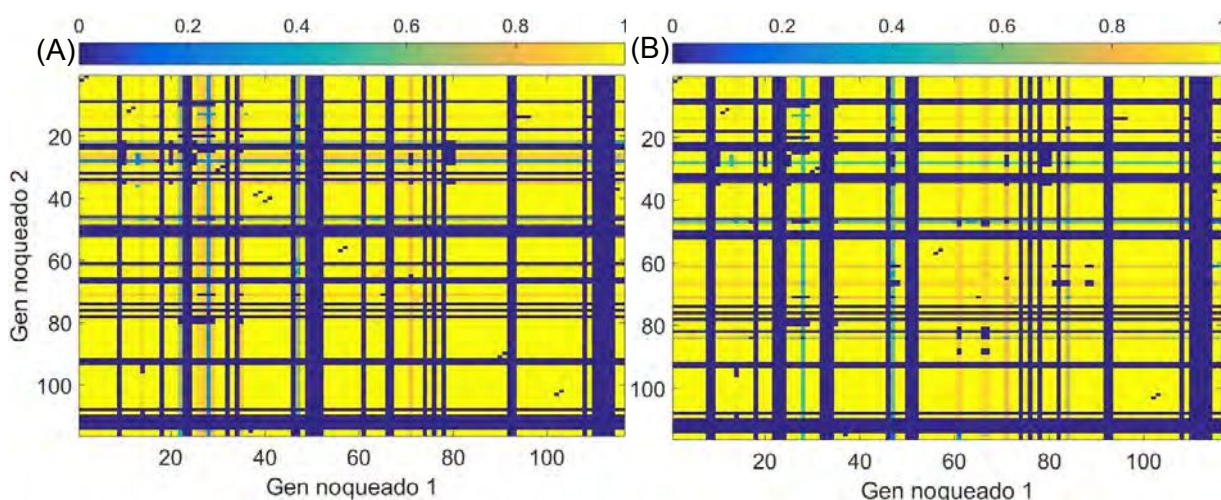


Fig. 2. Influencia del doble knockout de genes en el crecimiento de *R. toruloides* en glucosa (A) y glicerol (B) como FCE.

3.2.- Simulación de doble knockout de genes.

Para el análisis de genes redundantes se estudió el efecto del knockout simultáneo de dos genes sobre el crecimiento de *R. toruloides* para glucosa y alternativamente para glicerol (Fig. 2). En estos gráficos, la escala de colores representa el *grRatio*, con lo cual el color azul simboliza genes esenciales para el crecimiento, mientras que el color amarillo son genes no esenciales. Los colores verde y naranja representan genes cuyos knockouts implican la reducción de la tasa de crecimiento del microorganismo, pero no su letalidad. En lo que respecta a genes esenciales, las líneas de color azul intenso representan los genes que son esenciales por sí mismos, o sea son los 25 genes esenciales que se determinaron mediante el knockout individual de genes. Por otro lado, las zonas puntuales de color azul representan a genes redundantes, esto es, son genes que individualmente no son esenciales pero el knockout combinado con otro gen resulta letal. Para *R. toruloides* creciendo en glucosa como única FCE se encontraron 59 genes esenciales redundantes mientras que con glicerol se hallaron 75 genes esenciales redundantes. El conocimiento de estos genes es fundamental si la estrategia de sobreproducción lipídica incluye el knockout de más de un gen a la vez.

En cuanto al análisis de esencialidad de los genes esenciales redundantes, se determinó la existencia de 48 genes redundantes esenciales mínimos, lo que implica que independientemente de la FCE, el doble knockout de estos genes produce la letalidad de la cepa. Además, con glucosa y glicerol como FCE hay 11 y 27 combinaciones letales adicionales, respectivamente, constituyendo así genes redundantes condicionales a la FCE. Dada la existencia de genes condicionales al entorno resulta fundamental tener en cuenta las características del medio y condiciones de cultivo a la hora de evaluar *in silico* la esencialidad de genes, previo a la realización de ensayos experimentales.

4.-CONCLUSIONES

En conclusión, esta herramienta bioinformática permitió determinar de forma rápida y sencilla el conjunto de genes cuyo knockout resulta individualmente, o en combinación, letal para el microorganismo. Esto permitirá ahorrar tiempo y recursos a la hora de diseñar mutantes para la sobreproducción de lípidos microbianos.

Futuros trabajos estarán dirigidos a ampliar el modelo metabólico para tener más información acerca de posibles blancos para la modificación genética de la cepa, así como su eventual validación experimental.

5.-AGRADECIMIENTOS

Este trabajo fue financiado por proyectos de la Agencia Nacional de Promoción Científica y Tecnológica (PICT 2014-2394 y PICT 2016-2258), CONICET (PIP 2015-0837) y la Universidad Nacional de La Plata (Proyecto I216, PPID-I008).

6.-REFERENCIAS

- [1] Soccol, C. R *et al.* (2017). Pilot scale biodiesel production from microbial oil of *Rhodospiridium toruloides* DEBB 5533 using sugarcane juice: performance in diesel engine and preliminary economic study. *Bioresource Technology*, 223, 259-268.
- [2] Faife-Pérez, E *et al.* (2012). Producción de biodiesel a partir de microorganismos oleaginosos. Una fuente de energía renovable. Parte I. Levaduras y bacterias. *ICIDCA. Sobre los Derivados de la Caña de Azúcar*, 46(1):22-32.
- [3] Papanikolaou, S., & Aggelis, G. (2011). Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. *European Journal of Lipid Science and Technology*, 113(8), 1031-1051.
- [4] Ling, J., Nip, S., & Shim, H. (2013). Enhancement of lipid productivity of *Rhodospiridium toruloides* in distillery wastewater by increasing cell density. *Bioresource Technology*, 146, 301-309.
- [5] Fei, Q. *et al.* (2016). Enhanced lipid production by *Rhodospiridium toruloides* using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. *Biotechnology for Biofuels*, 9(1), 130.
- [6] Castañeda, M. T *et al.* (2018). Comprehensive analysis of a metabolic model for lipid production in *Rhodospiridium toruloides*. *Journal of Biotechnology*, 280, 11-18.
- [7] Yang, X *et al.* (2014). Recycling biodiesel-derived glycerol by the oleaginous yeast *Rhodospiridium toruloides* Y4 through the two-stage lipid production process. *Biochemical Engineering Journal*, 91, 86-91.
- [8] Wu, S *et al.* (2011). Microbial lipid production by *Rhodospiridium toruloides* under sulfate-limited conditions. *Bioresource Technology*, 102(2), 1803-1807.
- [9] Zhao, X *et al.* (2011). Lipid production by *Rhodospiridium toruloides* Y4 using different substrate feeding strategies. *Journal of Industrial Microbiology & Biotechnology*, 38(5), 627-632.
- [10] Yang, X *et al.* (2018). Expression of phosphotransacetylase in *Rhodospiridium toruloides* leading to improved cell growth and lipid production. *RSC Advances*, 8(43), 24673-24678.
- [11] Zhang, S *et al.* (2016). Engineering *Rhodospiridium toruloides* for increased lipid production. *Biotechnology and Bioengineering*, 113(5), 1056-1066.
- [12] Lin, X *et al.* (2014). Functional integration of multiple genes into the genome of the oleaginous yeast *Rhodospiridium toruloides*. *FEMS Yeast Research*, 14(4), 547-555.
- [13] Long, M. R., Ong, W. K., & Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Current Opinion in Biotechnology*, 34, 135-141.
- [14] Orth, J. D., Fleming, R. M., & Palsson, B. O. (2010). Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. *EcoSal plus*.
- [15] Xu, J., Zhao, X., Wang, W., Du, W., & Liu, D. (2012). Microbial conversion of biodiesel byproduct glycerol to triacylglycerols by oleaginous yeast *Rhodospiridium toruloides* and the individual effect of some impurities on lipid production. *Biochemical Engineering Journal*, 65, 30-36.
- [16] Apweiler, R *et al.* (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 32(suppl_1), D115-D119.
- [17] Llaneras, F., & Picó, J. (2008). Stoichiometric modelling of cell metabolism. *Journal of Bioscience and Bioengineering*, 105(1), 1-11.
- [18] Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245.
- [19] Palsson, B. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge: Cambridge University Press.
- [20] Schellenberger, J *et al.* (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature protocols*, 6(9), 1290.
- [21] Zhang, Z., & Ren, Q. (2015). Why are essential genes essential?-The essentiality of *Saccharomyces* genes. *Microbial Cell*, 2(8), 280.